

## **Exhibit 2**

**IN THE UNITED STATES DISTRICT COURT  
FOR THE EASTERN DISTRICT OF TEXAS  
SHERMAN DIVISION**

The State of Texas, et al.,

*Plaintiffs,*

v.

Google LLC,

*Defendant*

Civil Action No. 4:20-cv-00957-SDJ

**DECLARATION OF JACOB HOCHSTETLER IN SUPPORT  
OF PLAINTIFF STATES' MOTION FOR SPOILATION SANCTIONS**

1. My name is Jacob Hochstetler and I am a Clinical Assistant Professor of Computer Science and Engineering at the University of North Texas. I hold B.S. and M.S. degrees in computer science and a Ph.D. in computer science and engineering from the University of North Texas. I have professional experience analyzing logs for high performance computing applications. My CV is attached to this declaration as Exhibit 2-A.

2. As part of my assignment in this case, I analyzed backend system-wide logs (the “Log Dataset”) <sup>1</sup> that Google produced to the Plaintiff States containing information regarding Google employees’ use of its instant messaging platform (“Google Chat”), including entries detailing the times that each employee sent Google Chat messages during the period and whether the employee sent each Google Chat message with history setting “on” or “off.”<sup>2</sup> Google’s log

---

<sup>1</sup> The Log Dataset consists of the following files: GOOG-AT-MDL-C-000088198.CSV, GOOG-AT-MDL-C-000088199.CSV, GOOG-AT-MDL-C-000088200.CSV, GOOG-AT-MDL-C-000088201.CSV, GOOG-ATMDL-C-000088202.CSV, GOOG-AT-MDL-C-000088203.CSV, GOOG-AT-MDL-C-000088204.CSV, GOOG-AT-MDL-C-000088205.CSV, GOOG-AT-MDL-C-000088206.CSV, GOOG-AT-MDL-C-000088207.CSV, GOOG-AT-MDL-C-000088208.CSV, GOOG-AT-MDL-C-000088209.CSV, GOOG-ATMDL-C-000088210.CSV, GOOG-AT-MDL-C-000088211.CSV.

<sup>2</sup> See Lopez Dep. 218:19-222:3.

production to the States included five Google employees whom I understand were designated as document custodians in this case including Google's CEO Sundar Pichai. The Log Dataset spans a period of 68 days from December 9, 2022 to February 14, 2023 (the "Log Period").<sup>3</sup>

3. My analysis of the Log Dataset confirmed that Google's systems did not automatically set the default Google Chat history setting for a Google Chat conversation to "on" for the five custodians' messages until February 8, 2023. Instead, prior to February 8, 2023, messages in the Log Dataset defaulted to Google Chat history setting "off." On February 8, 2023, messages in the Log Dataset began defaulting to Google Chat history setting "on." This is consistent with my understanding that Google changed the default Google Chat history setting for employees on litigation hold starting in February 2023.

4. Based on the testimony of Google's corporate representative [REDACTED], any Google Chat messages sent with the Google Chat history "off" were never available to be retained by Google Vault and were likely permanently deleted 24 hours after being sent.<sup>4</sup> Thus, any Google Chat messages sent with the Google Chat history setting "off" in 2022 were likely deleted.

5. My analysis showed that at no point during the Log Period did any of the five custodians represented in the Log Dataset personally change the Google Chat history setting from "off" to "on" for any of the conversations in which they participated during the Log Period.

6. Based on my analysis of the information in the Log Dataset, those five custodians sent a total of approximately 21,269 Google Chat messages during the Log Period.

---

<sup>3</sup> The Log Period of December 9, 2022 to February 14, 2023 includes a period in which Google's corporate Google Chat history default setting was "off" (December 9, 2022–February 8, 2023) and a period in which Google's corporate Google Chat history default setting was "on" (February 8, 2023–February 14, 2023).

<sup>4</sup> Lopez Dep. 18:21-19:10, 75:24-76:1 77:7-78:19, 90:2-9,136:6-13, 242:13-18, 244:11-22

7. Based on my analysis of the information in the Log Dataset, 87% of those 21,269 Google Chat messages were sent with the Google Chat history setting “off.” Specifically, I determined that at least 18,566 out of the approximately 21,269 Google Chat messages those five custodians sent during the Log Period were sent with the Google Chat history setting “off.”

8. Based on my analysis of the information in the Log Dataset, aggregated across the five Google custodians, 94.5% of Google Chat conversations reflected in the Log Dataset had the Google Chat history setting turned “off” for at least some time during the Log Period prior to Google’s February 8, 2023 change to its corporate default Google Chat history setting.

9. Based on my analysis of the information in the Log Dataset, during the Log Period prior to February 8, 2023, 96% of all Google Chat messages involving Mr. Pichai in the Log Dataset were sent with the Google Chat history setting “off.”

10. My analysis also showed that Google Chat conversations which Google produced in this litigation that at least partially occurred during the Log Period had the Google Chat history “off” at other points during the Log Period. As one illustrative example, Google produced a portion of a Chat conversation involving Mr. Pichai to the Plaintiff States at GOOG-AT-MDL-007412389-93. That Chat conversation was reflected in the Log Dataset. The Log Dataset contains records of all the Chat messages sent or received in this conversation during the Log Period. Approximately 387 Google Chat messages were sent or received in this Google Chat conversation during the Log Period. Of those 387 total Chat message, 157 were sent messages and approximately 230 were received messages.<sup>5</sup> Yet the conversation Google produced at GOOG-AT-MDL-007412389-93

---

<sup>5</sup> These numbers were obtained by filtering within the “pichai” tab of the “all\_retention\_states.xlsx” spreadsheet produced with my Supplemental Report. Applying the filters “space\_id” = “AAAAbOkmVBe” and “action” = “DynamiteChimeRendererService-Render” and counting the results yields 691 total received messages. As received message logs are

only contains 61 total messages.<sup>6</sup> Based on my analysis of the Log Dataset, I have confirmed that there were approximately 135 sent messages and 199 received messages with the Google Chat history setting “off.” In other words, at least 86% of the Google Chat messages in that chat conversation were likely deleted 24 hours after being sent and would have been unable to be retained or produced.

11. Based on the 18,566 messages<sup>7</sup> sent by the five custodians represented in the Log Dataset with Google Chat history “off” over the 68-day Log Period, I conclude that a given Google custodian likely sent approximately 20,000 Google Chat messages with the Google Chat history setting “off” during 2022.<sup>8</sup>

12. My estimate is conservative because the 68-day Log Period, on which that conclusion is based, included end-of-year holidays, and because one of the five individual custodians had extremely low, sporadic messaging volume.

13. In August 2024, Google represented to the States that it had placed 141 Google employees on a litigation hold for this matter.<sup>9</sup> Multiplying the likely number of messages sent with Google Chat history setting “off” per year by a Google custodian by the total number of

---

often duplicated three times (but occasionally duplicated fewer times), dividing by 3 yields an (under)count of approximately 230 received messages. Switching the “action” filter to “action” = “DynamiteService-CreateTopic” and “DynamiteService-CreateMessage” and counting the results yields 157 total sent messages. Adding 230 and 157, which equals 387, yields an approximation of the total chat messages in the conversation according to the Log Dataset.

<sup>6</sup> There were 61 messages produced in this Google Chat conversation, counted via a visual inspection.

<sup>7</sup> See *supra* ¶3.

<sup>8</sup> As I detail in my Supplemental Report, FN 4, this number is based on an estimated 18,566 messages lost for the 5 individuals represented in the Log Dataset over 68 days. Hence, I found that  $18,566/(5 \times 68) \times 365 = 19,931$  messages (which I have rounded to 20,000) lost per employee in a given year due to a message being sent with Google Chat history setting “off” prior to the change of the Retention Setting default is a conservative estimate for a given employee.

<sup>9</sup> Letter from Robert McCallum, August 29, 2024

Google custodians in this matter, I conclude that it is likely that those 141 custodians sent approximately 2.8 million Google Chat messages with the Google Chat history setting “off” during 2022.

14. If instead I assume that, due to turnover or other reasons, half of those 141 custodians under litigation hold in this matter had left Google by 2022, then, using the same formula, I conclude that it is likely that those 70 custodians sent approximately 1.4 million Google Chat messages with the Google Chat history setting “off” during 2022.

15. I also understand that Google’s expert Michal Malkiewicz has asserted that Google has actually placed 202 employees under a litigation hold for this matter. Taking that as true, I conclude that the total number of Google Chat messages sent with Google Chat history “off” during 2022 by Google’s custodians in this matter is actually higher. Using that litigation hold count and applying the same formula, I conclude that it is likely that those 202 custodians sent approximately 4 million Google Chat messages with the Google Chat history setting “off” during 2022.

16. I have seen no evidence that suggests that Google’s employees under a litigation hold would send fewer chats with the Google Chat history setting “off” in any year during the relevant period.

17. I have seen no evidence that suggests that the number of Google Chat messages sent and received by relevant Google employees under a litigation hold would have been lower in any other years during the relevant period than in 2022.

18. I also understand that Google returned to office from the pandemic starting in April 2022. As a result, I would expect that Google employees would have sent at least as many Google Chat messages in the years from 2019 through 2021 as they did in 2022, as remote work would

likely increase electronic communications. Further, as a technology company and the inventor of Google Chat, it is reasonable to expect that Google employees used the Google Chat platform at similar rates to years prior to that as well.

19. My review of both the Log Dataset and the deposition testimony of Google employees and corporate representatives supports my conclusions. That review indicates that Google employees generally used Google Chat every day.<sup>10</sup>

20. I have seen no evidence that suggests that the number of Google custodians that should have been on a litigation hold during the relevant period before 2022 was lower than 202.

21. I have seen no evidence that suggests that the five custodians use of Google Chat messages reflected in the Log Dataset—specifically the sending and receiving of Google Chat message with the Google Chat history setting “off”—is not reflective of the way other Google employees under a litigation hold in this case used Google Chat or the Google Chat history setting.

22. During the course of this litigation, Google has produced approximately 4.2 million emails.<sup>11</sup> Those emails were sent in the following years of the relevant period: <sup>12</sup>

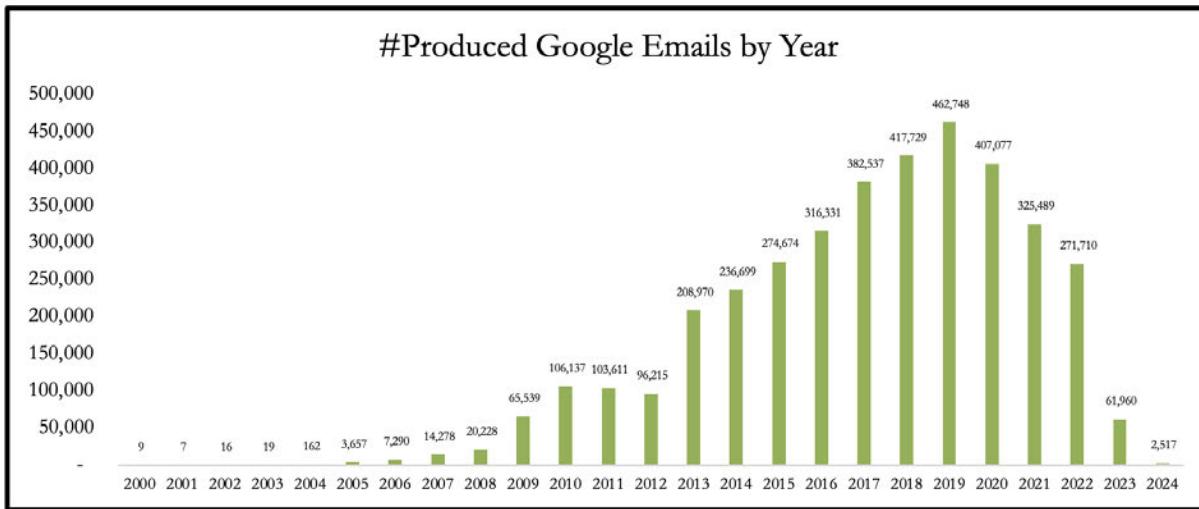
---

<sup>10</sup> Lopez Dep. 228:3-229:6, “Q. Okay. Let’s go to paragraph 14 on page 6. The court found that, the first sentence, Google chat is an essential tool used daily by Google employees. Do you have any reason to disagree with that? ... A. In my personal experience, that’s correct. Q. Okay. Paragraph 15. The Court found: There are no restrictions on the content and topics of Chat, ... Chat can be used for, quote, anything under the sun that employees want to communicate... Is that accurate? ... A. Yes, the product does not restrict you from typing anything that you might be interested in communicating.” *Also see* Pichai Dep. 33:6-11, “Q. All right. Now, is it also true that Google used Chat for purposes including anything sensitive? A. We definitely use Chat for regular communication as part of our ongoing work.”

<sup>11</sup> I identified ~4.2 million emails from the documents produced by Google by filtering for documents with email extensions (.eml,.msg, msg, eml,.mht, and email). I used the date-time meta data associated with these documents to identify which year the email document corresponded to.

<sup>12</sup> Out of the ~4.2 million emails produced by Google, 400,524 emails do not have date-time metadata associated with them. So, I have included only ~3.8mn emails in the plot “#Produced Google Emails by Year” where the email counts by year is plotted. Out of the ~3.8mn emails 7,490 were flagged as exact duplicates and ~1.8mn were flagged as near duplicates. This may potentially

**Figure 1**



23. During the course of this litigation, Google has produced approximately 14-15 thousand Google Chat conversations.<sup>13</sup> Those Google Chat conversations were sent in the following years of the relevant period and I have not identified a single chat produced prior to 2010 based on my review:<sup>14</sup>

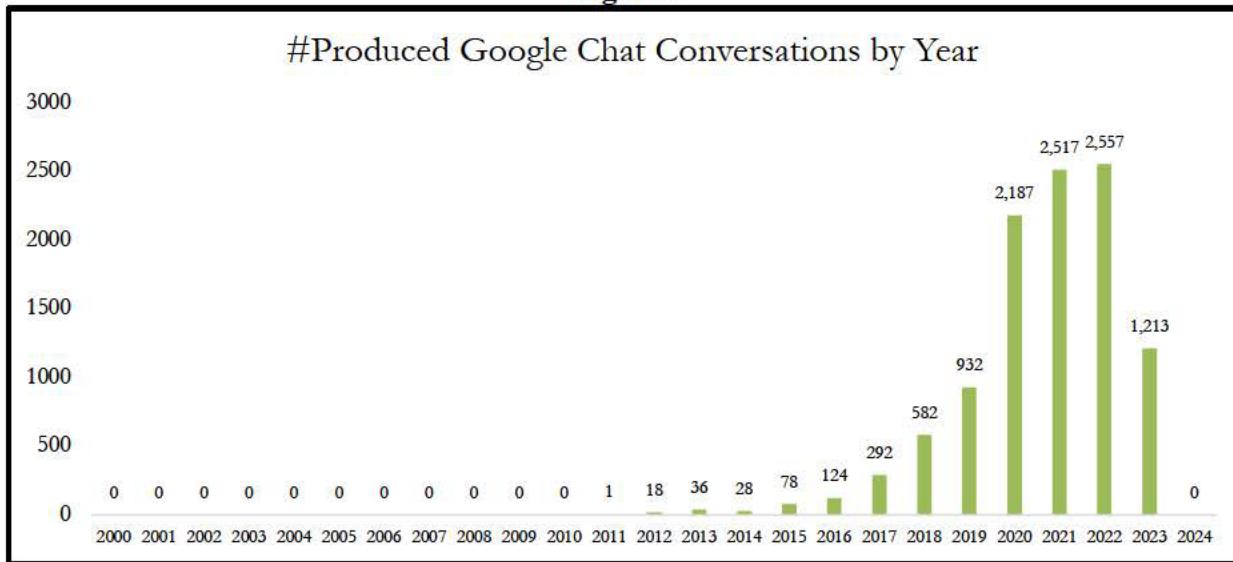
---

be due to different emails from the same email thread produced with different Bates numbers. I have not performed any deduplication of the email documents.

<sup>13</sup> I identified ~15k chat conversations from the documents produced by Google by filtering for documents tagged “google chat”, “\_google\_chat\_”, “gmail chat”, or “\_gmail\_chat\_” or contained “AAAA” or “threaded” in the subject. I have also included 13,334 documents identified by Google as chat conversations (See State of Texas v Google LLC\_4-20-cv-00957-SDJ\_Bates List.xlsx). I used the date-time meta data associated with these documents to identify which year the email document corresponded to.

<sup>14</sup> Out of the ~14-15k chat conversations produced by Google, 4,225 chats do not have date-time metadata available. Thus, I have included only ~9k chats in the plot “#Produced Google Chats Conversations by Year” where the total chat conversation counts by year is plotted. Out of the ~9k chats, 61 were flagged as exact duplicates and 335 were flagged as near duplicates. This may potentially be due to different chat conversations from the same chat produced with different Bates numbers. I have not performed any deduplication of the chat conversation documents.

**Figure 2**



24. Based on my review of the Log Dataset, the depositions of Google employees, and the database of documents produced by Google, the year that the most produced Google emails were created/sent is different than the year that the most produced Google chats were sent/received.

I declare under penalty of perjury the laws of the United States of America that the foregoing is true and correct. Executed this December 8, 2024, in Roanoke, Texas.



Jacob Hochstetler, Ph.D.

## **Exhibit 2-A**

# JACOB HOCHSTETLER, PhD

(940) 765.9451 | [jacob.hochstetler@gmail.com](mailto:jacob.hochstetler@gmail.com) | [GitHub Profile](#) | [LinkedIn](#) | [ResearchGate](#)

## Education

**Ph.D.** – Computer Science and Engineering – University of North Texas, Denton, TX – 2021

Dissertation: An Extensible Computing Architecture Design for Connected Autonomous Vehicle System

Advisor: Dr. Song Fu

**Master of Science** – Computer Science – UNT – 2018

**Bachelor of Science** – Computer Science – UNT – 2011

## Publications

- *Cooperative Mixed Reality Leveraging Edge Computing and Communication* – 5<sup>th</sup> ACM/IEEE Symposium on Edge Computing (SEC) – 2020
- *Low-latency High-level Data Sharing for Connected and Autonomous Vehicular Networks* – IEEE Intl. Conf. on Industrial Internet (ICII) – 2019
- *An Empirical Study of Quad-Level Cell (QLC) NAND Flash SSDs for Big Data Applications* – IEEE Intl. Conf. on Big Data (Big Data) – 2019
- *Embedded Deep Learning for Vehicular Edge Computing* – 3<sup>rd</sup> IEEE/ACM Symposium on Edge Computing (SEC) – 2018
- *Reliability Characterization of Solid State Drives in a Scalable Production Datacenter* – IEEE Intl. Conf. on Big Data (Big Data) – 2018
- *Incorporate Proactive Data Protection in ZFS Towards Reliable Storage Systems* – IEEE 16<sup>th</sup> Intl. Conf. on Dependable Computing (DASC) – 2018
- *Developing Cost-Effective Data Rescue Schemes to Tackle Disk Failures in Data Centers* – Springer Intl. Conf. on Big Data – 2018
- *An Optimal Police Patrol Planning Strategy for Smart City Safety* – IEEE 14<sup>th</sup> Intl. Conf. on Smart City (SmartCity) – 2016

## Clearances & Certifications

- **TS/SCI** – DoD – 1997
- **Counter-intelligence polygraph** – DoD – 1997
- **Full-scope polygraph** – CIA – 2003
- **Security+** – CompTIA – 2010
- **ITILv3 Certified** – AXELOS – 2010
- **Sun Certified System Administrator (SCSA)** – Sun Microsystems – Solaris 8 & 2.6
- **Sun Certified System Support Engineer (SSE)** – Sun Microsystems – 2002
- **Cisco Certified Network Associate (CCNA)** – Cisco – 2000

## Litigation Support & Expert Witness Experience

**The State of Texas v. Google LLC** – TX OAG/Keystone Strategy – TX Eastern District, 4:20-cv-00957

2024 to present

- **Matter:** Antitrust, Online ad tech and monopolization
- **Provided:** Expert witness: code review, opening technical report, rebuttal report, supplemental report, deposition

**Cellspin Soft, Inc. v. Fitbit, Inc., et al.** – Garteiser Honea – CA Northern District, 4:17-cv-05928

2020 to 2023

- **Matter:** Patent Infringement, Digital Data capture with Bluetooth interface
- **Provided:** Seven code reviews

**InfoGation Corp. v. Google LLC** – Sheridan Ross PC – TX Western District 6:20-cv-0366

2020 to 2022

- **Matter:** Patent Infringement, Mobile Navigation System
- **Provided:** Code review

**Tactical Entertainment, LLC v. Krasamo, Inc.** – Creedon PLLC – TX 401<sup>st</sup> District, 401-03246-2017

2019 to 2021

- **Matter:** Software development, Back-end services, APIs, and Networking
- **Provided:** Code review

## Teaching Experience

---

### Clinical Assistant Professor, University of North Texas, Denton, TX

- CSCE 1015 – *Computing Tools and Techniques Laboratory*: Hands-on introduction to essential tools and techniques for computer science.
- CSCE 3055 – *IT Project Management*: Microsoft Project, SDLC, Agile development, Git, Basecamp, Project Finance, Google Analytics.
- CSCE 3220 – *Human Computer Interfaces*: UI/UX, Adobe XD, Web Accessibility, iOS UI and Android UI kits.
- CSCE 3420 – *Internet Programming*: PHP, JavaScript, Node.js, HTML/CSS, Client/Server architecture, APIs, AWS Lambda.
- CSCE 3530 – *Introduction to Computer Networks*: OSI Model: MAC, Ethernet, TCP/IP, Proxies/Load Balancers up to Application layer.
- CSCE 3550 – *Foundations of Cybersecurity*: Security goals, threats, vulnerabilities. Network, program, and operating system security issues.
- CSCE 4350 – *Database Systems*: SQL, MySQL, Object Stores, NoSQL, AWS DynamoDB.
- CSCE 4560 – *Secure Electronic Commerce*: Shopping carts/Payment gateways, eCommerce security, PKI, SSL/TLS, authZ/authN, blockchain.
- CSCE 4600 – *Operating Systems*: System abstraction/virtualization, process/threads concurrency, persistence (resource management), security.
- CSCE 5350 – *Fundamentals of Database Systems*: Relational databases: design, querying, SQL, DBMS features, normalization, and PL/SQL.
- CSCE 5520 – *Wireless Networks and Protocols*: Wireless networks, protocols, encoding techniques, cellular networks, and LTE-Advanced.
- CSCE 5552 – *Cybersecurity Essentials*: Data concealment/obfuscation, system identity/reconnaissance/exploitation, cryptography, data forensics.
- CSCE 5580 – *Computer Networks*: Protocols, routing, topology, coding, and multiplexing in computer network communication.
- CSCE 5585 – *Advanced Network Security*: Firewalls, intrusion prevention/detection systems, network forensics, network pentesting.

## Research Experience

---

### University of North Texas, Dependable Computing Systems Lab, Denton, TX

- Data format and protocol end-to-end architecture for connected vehicle data sharing.
- Embedded/Single board computer edge node clustering using Rancher k3s and k3os.
- Autonomous, self-driving system using Nvidia PX2 with Lidar, Radar, and image sensors on a Polaris GEM (with NSF).
- Edge node-style hard-drive deployments using single-board computers embedded in HDD PCBs (in partnership with HP Labs).
- Real-time machine learning with Microsoft HoloLens for augmented reality and a single-board computer for inference.

### Los Alamos National Labs, Department of Energy, Trinity, Los Alamos, NM

- Machine learning log HPC analysis.
- Singularity container system to deploy HPC workloads.
- Analyzed millions of daily HPC log entries to detect and interpret anomalies.

## Work History

---

### Vice President, Personal Investing, Cloud Engineering [Distinguished Engineer] – Fidelity Investments – Westlake, TX

2022 to Present

- Provided technical oversight to Fidelity's Personal Investing (PI) Platform Engineering teams as a Distinguished Engineer.
- Created app-team level self-service routing based on multiple Envoy routers, both through public and on-premise datacenters.
- Developed "next-gen" Golang skeleton/framework for high-performance (high-TPS) applications.
- Facilitated high-level and cross-business unit architecture design teams.

### Clinical Assistant Professor, Department of Computer Science and Engineering – University of North Texas – Denton, TX

2021 to Present

- Taught undergrad and graduate level computer science, information technology, and cybersecurity courses.
- Developed online curriculum for multiple classes.
- Led NSA/NSF GenCyber student and teacher summer camps [2022, 2023, 2024].
- Chaired multiple hiring committees, resulting in onboarding twelve faculty members.
- Served as Computer Science curriculum program advisor.

**Director, Infrastructure as a Service Development/Product Owner/IC** – Fidelity Investments – Westlake, TX

2018 to 2022

- Led a “two-pizza team” developing an on-premise cloud management platform with Kanban-style feature delivery.
- Acted as lead individual contributor to application architecture, design, and development.
- Mentored junior developers on best practices in SDLC, CI/CD, and infrastructure automation/orchestration.
- Led stability project in migrating a decade-old legacy Python codebase to Golang.

**Principal Software Engineer/Cloud Technologist** – Fidelity Investments – Westlake, TX

2016 to 2018

- Maintained API (Python) and UI (Rails/AngularJS) for an on-premise cloud management platform.
- Provided DevOps support for both the cloud management platform and underlying infrastructure (Citrix XenServer, OpenStack, and AWS EC2).
- Developed “Day 0”, “1” and “2” automation through Chef and Ansible to meet security, compliance and audit requirements.

**Senior Software Engineer** – Fidelity Investments – Westlake, TX

2015 to 2016

- Led project to automate end-user SSH PKI Lifecycle Management within Fidelity.
- Developed highly available architecture composed of master-master DB nodes surrounded by “dumb” API and UI nodes.
- Used fpm to create packages for Solaris, Red Hat, Debian (Ubuntu), and AIX for seamless user integration (opensshd/sshd) with the PKI portal.

**UNIX Systems Lead** – General Dynamics – SW Asia

2011 to 2014

- Maintained UNIX systems for the US Air Force Central Command Intelligence, Surveillance and Reconnaissance Division.
- Provided Subject Matter Expertise to A6/G6/J6 commands, along with down-range assets.
- Developed real-time, self-service web services to replace business processes for the Change-Advisory Board, increasing delivery to customers.

**JWICS Systems/C2 Systems Lead** – General Dynamics – SW Asia/HOA/AFG/IRQ

2010 to 2011

- Supported core JWICS hardware/software including VMware, Solaris, NetApp filers, Windows, and Red Hat Enterprise Linux (RHEL).
- Maintained multiple JSTARS Workstations, Approver for Radiant Mercury in training and live environments.

**Sun Engineer (Onsite)** – General Dynamics – Plano, TX

2007 to 2010

- Provided on-site support to the EDS (later HP) datacenter for Sun Microsystems/Hewlett Packard/General Motors/OnStar.
- Developed in-house application to manage personnel rotations and inventory at all Sun-managed datacenter to meet SLAs.

**C2 Systems Administrator** – General Dynamics – SW Asia

2006 to 2007

- Installed, configured and maintained Solaris-based Command and Control systems at the Combined Air Operations Center.
- Trained end-users and maintainers on JSWS, Radiant Mercury, and Information Support Server Environment guard.
- Created and maintained Continuity of Operations (COOP) plans and documentation.

**GCCS Administrator/COP Manager** – General Dynamics – SW Asia

2005 to 2006

- Provided on-site operational and functional support as the Common Operational Picture manager at the Combined Air Operations Center.
- Managed the flow of information between the SIPR, CSV, CENTRIX, GCC+2 networks, Administered DII-COE GCCS cluster on Solaris 2.5.1.
- Manipulated displays and ground, air and sea tracks for local C2PC clients and downstream feeds.

**Sun Microsystems Subject Matter Expert Engineer (Onsite SME)** – General Dynamics – SW Asia/AFG/IRQ

2004 to 2005

- Conducted hardware and software fault isolation, root cause analysis and repair on Sun Microsystems servers, storage arrays, and workstations.
- Provided continual training to military personnel in the operation of Solaris and Sun Microsystems equipment.

**Security Contractor** – Academi (formerly Blackwater Security Company) – SW Asia/IRQ

2003 to 2004

- Conducted personal security detail and training operations for the US and Iraqi military.

**Field Engineer** – General Dynamics – San Jose, CA

2000 to 2002

- Serviced Sun Microsystems equipment at secure (GOV/MIL) and commercial customer sites.
- Developed in-house reporting tools to gather email and SMS messages, parse them for job information, and assign escalation queues.

**Counter-Intelligence Agent – US Army**

1996 to 2005

- Provided CI duties as a human intelligence (HUMINT) agent to sliced elements.
- Surveilled, interviewed, and interrogated human sources, and analyzed open-source (OSINT) intelligence.
- Engaged in security/force-protection operations involving intelligence collection, processing, producing, and dissemination.